



TITLE:

複雑ネットワークの構造特性と構成要素の属性分布の関係(第2回生物数学の理論とその応用)

AUTHOR(S):

守田, 智

CITATION:

守田, 智. 複雑ネットワークの構造特性と構成要素の属性分布の関係(第2回生物数学の理論とその応用). 数理解析研究所講究録 2006, 1499: 24-29

ISSUE DATE:

2006-07

URL:

<http://hdl.handle.net/2433/58402>

RIGHT:

複雑ネットワークの構造特性と構成要素の属性分布の関係

静岡大学 工学部 システム工学科 守田 智 (Satoru Morita)
Department of Systems Engineering, Shizuoka University

1 はじめに

生物の代謝系, 生態系から社会システムや工学的通信システムに至るまで複雑なシステムの構造・機能を解析するためには, そのネットワーク構造に着目することが必要である. 多くのシステムのネットワークは正方格子のように規則的でもないし, ランダムグラフのようにでたらめでもなく複雑ネットワークと呼ばれる範疇に属している. 複雑ネットワークのよく知られている特徴としてスモールワールド性とよばれる性質がある [1]. スモールワールド性とは, クラスタリング係数 C が大きいにもかかわらず, ランダムグラフ並みにノード間の平均最短距離 L が短いということをいう. ここでクラスタリング係数とは, あるノードに着目してそのノードとつながるノードを2つ選択したときにその2つのノード間にもリンクが存在する割合である. 一般的にはすべてのノードに対する平均値が用いられることが多い. また複雑ネットワークの数理モデルの多くで平均ノード間距離 L がノード数 N の増加に対して対数 $L \propto \ln N$ かそれより遅い増加を示す [2, 3, 4]. このような遅い増加を示す場合に平均ノード間距離が小さいといえる. さらに複雑ネットワークの多くは, スケールフリーという特徴も兼ねそなえている. これは, 結合次数の分布が

$$P(k) \sim k^{-\nu}$$

という冪則に従っていることを意味しており, ほとんどの場合その指数が $2 < \nu < 3$ を満たしていることが知られている [2, 3, 5, 6, 7, 8].

本研究の目的は, ネットワークを構成するノードが属性によって特徴付けられると場合, その属性がネットワークの特性に与える影響を解析することである. このような解析によってネットワークがどのような特性を持ちうるかを検討できる. 単純なモデルとしてノードの属性が「位置」 x_i と「結合の生じやすさ」 a_i の2つで表せる場合を考える. 「位置」はノードを空間上に配置することで定義され, ノードのペアに結合があるかどうかは空間上の距離の影響を受ける. すなわち, 近ければ近いほど繋がりやすいとする. また「結合の生じやすさ」はこの指数が大きいノードほど結合が生じる確率が高いことを表す. これはノード自体がもつ勢力の強さに対応する. ここでは結合の生じやすさの分布が冪則 $\rho(a) \propto a^{-\gamma}$ に従うと仮定する. 遺伝子の発現量, 都市の人口の分布, 企業のサイズなどの多くの量でこのような冪則が観測されており γ がおよそ2の場合に zipf 則と呼ばれている. したがって冪則の仮定は, 現実的であるといえる. このようなモデルを解析することによりノードの属性値 a の冪指数 γ とネットワークの次数 k の冪指数の関係を明らかにし, 空間構造 (ノードの位置) の導入によってスモールワールド性がどう再現されるかを見る.

2 モデル

まず N 個のノードを考える. i 番目のノードは「結合の生じやすさ」を示す属性値 a_i を持つとする. ここで a_i は

$$a_i = \left(\frac{i}{N}\right)^{\frac{1}{1-\gamma}} \quad (i = 1, 2, \dots, N) \quad (1)$$

というふうに決定論的に与える. ただし $\gamma > 1$ とする. 指数が $\gamma \simeq 2$ を満たす場合に Zipf 則に対応する. ノードの総数 N が十分大きいと仮定すると属性値 a_i の分布は

$$\rho(a) = (\gamma - 1)a^{-\gamma} + \frac{\delta(a - 1)}{2N} + \frac{\delta(a - N^{\frac{1}{\gamma-1}})}{2N} \quad (2)$$

と近似でき,

$$1 \leq a \leq N^{\frac{1}{\gamma-1}} \quad (3)$$

でのみ正の値を持つ. ここで $\delta(x)$ は Dirac のデルタ関数である. このように a の分布は両端を除くと冪則 $\rho(a) \propto a^{-\gamma}$ に従う.

また, 各ノードは d 次元立方体内の点であるとする. ノードの分布は一様で, 2つのノード i, j 間の距離 $l(i, j)$ は周期境界条件を課した最大ノルムで定義する. 2つのノード i, j に結合が生じる条件を

$$\frac{(2l(i, j))^d}{a_i a_j} < \theta \quad (4)$$

とする. θ は結合の総本数を決める閾値パラメーターである. ここでは結合本数が mN となるように θ を選ぶ. 上式で d 乗しているのは結合の生じやすさ a が倍になれば結合が生じる範囲の d 次元体積が倍になるようにするためである. 上式の左辺の分母を結合の生じやすさの積にしているのも同様の理由からである. 上の式を和によって定義することもできるが適当な変換によって積の形に置き換えることができる [9]. その場合は a の分布は別のものとなる. 本論文では (4) で a を定義したものとする.

3 次数分布

ノードの空間上の分布が一様なのでノード間距離の分布は

$$p(l < x) = \begin{cases} (2x)^d & (l < 1/2) \\ 1 & (l \geq 1/2) \end{cases} \quad (5)$$

となる. したがって結合の生じやすさが a と a' で与えられているノードのペア間に結合がある確率は

$$r(a, a') = \min(\theta a a', 1) \quad (6)$$

である. これは確率は1を超えないためである. 結合の生じやすさが a であるノードの平均結合次数は

$$\bar{k}(a) = N \int r(a, a') \rho(a') da' \quad (7)$$

となり, (2) と (6) から

$$\bar{k}(a) \simeq \begin{cases} \frac{2(\gamma-1)N - \gamma N^{\frac{1}{\gamma-1}}}{2(\gamma-2)} \theta a & (a < \theta^{-1} N^{\frac{1}{\gamma-1}}) \\ \frac{(\gamma-1)\theta a - (\theta a)^{\gamma-1}}{\gamma-2} N & (a \geq \theta^{-1} N^{\frac{1}{\gamma-1}}) \end{cases} \quad (8)$$

と近似できる. さらに

$$2m = \int \bar{k}(a) \rho(a) da \quad (9)$$

となるので面倒な計算の後, θ の近似式

$$\theta \simeq \begin{cases} \frac{2m}{N} \left(\frac{\gamma-2}{\gamma-1} \right)^2 & (\gamma > 2) \\ \left(\frac{2m(2-\gamma)}{N \ln N} \right)^{\frac{1}{\gamma-1}} & (1 < \gamma < 2) \end{cases} \quad (10)$$

が得られる. このように漸近的な性質は $\gamma = 2$ を境に変化する. すなわち $\gamma > 2$ で (8) は N の大きい極限で

$$\bar{k}(a) \simeq 2m \frac{\gamma-2}{\gamma-1} a \quad (\gamma > 2). \quad (11)$$

となる一方, $1 < \gamma < 2$ では $a > \theta^{-1} N^{\frac{1}{\gamma-1}}$ の範囲で

$$\bar{k}(a) \simeq \frac{2m}{\ln N} a^{\gamma-1} \quad (1 < \gamma < 2) \quad (12)$$

と書き表せる. $1 < \gamma < 2$ の場合は結合次数は結合の生じやすさに比例せず, その $\gamma - 1$ 乗に比例する. このような違いは, (7) の積分の計算において $\gamma > 2$ では積分範囲のうち a が小さいほう ($a \sim 1$) が支配的である一方, $\gamma < 2$ では a が大きいほう ($a \sim N^{1/(\gamma-1)}$) が支配的となることから生じる.

結合次数の分布は

$$P(k) = \int P(k|a) \rho(a) da. \quad (13)$$

と書き表せる. ここで $P(k|a)$ は結合の生じやすさが a であるノードに対して次数が k である (条件付) 確率であり, 2 項分布

$$P(k|a) = \binom{N}{k} \left(\frac{\bar{k}(a)}{N} \right)^k \left(1 - \frac{\bar{k}(a)}{N} \right)^{N-k}. \quad (14)$$

で与えられる. $\gamma > 2$ の場合, (13) の積分の中身が最大値を取る a が積分範囲 (3) の内側にある条件

$$2m \frac{\gamma-2}{\gamma-1} + \gamma < k < 2m \frac{\gamma-2}{\gamma-1} N^{\frac{1}{\gamma-1}} + \gamma \quad (15)$$

をみたす k に対して鞍点法近似により

$$P(k) \simeq \frac{(2m)^{\gamma-1}}{N^{\gamma-1}} \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}} \frac{N!}{k!} \frac{\Gamma(k-\gamma+1)}{\Gamma(N-\gamma+2)}. \quad (16)$$

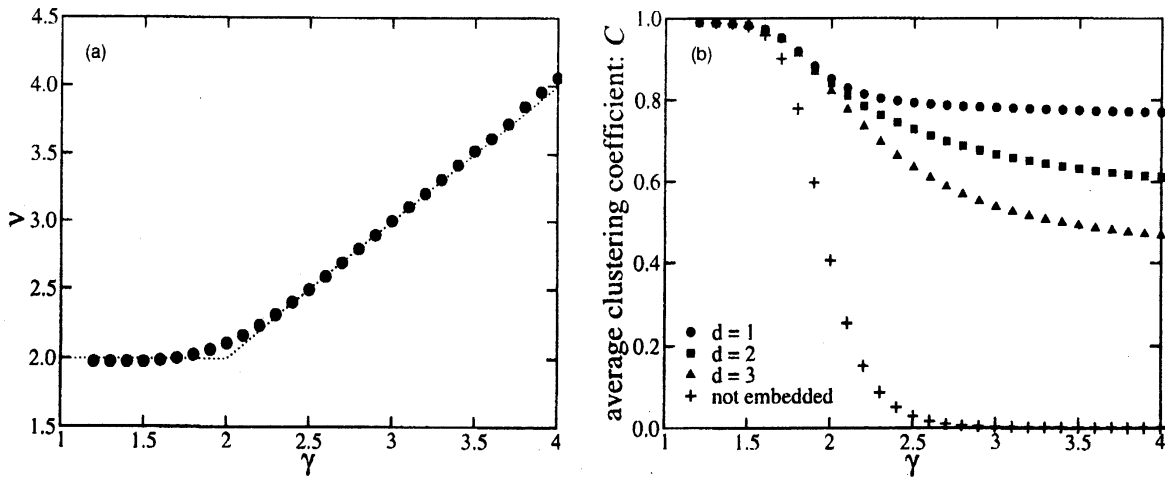


図 1: (a) 結合の生じやすさ a の分布の冪指数に対するネットワークの次数分布の冪指数. 点線は解析的な計算によって求めた予想 (17), (19) による. 黒丸は数値計算の結果であり, $m = 3, d = 2, N = 10000$ の場合である. この結果は d にはよらない. (b) 結合の生じやすさ a の分布の冪指数に対する平均クラスタリング係数. “not embedded” は空間にノードを埋め込まず, (6) の確率によってネットワークを構築したものである. $\gamma < 2$ ではクラスタリング係数は大きく次元 d による依存性は小さい. 一方, $\gamma > 2$ ではクラスタリング係数は次元 d が小さいほど大きくなり, 空間に埋め込まれてない場合にはほとんど 0 となっている.

となる. 特に $k \gg 1$ に対して

$$P(k) \simeq (2m)^{\gamma-1} \frac{(\gamma-2)^{\gamma-1}}{(\gamma-1)^{\gamma-2}} k^{-\gamma} \quad (\gamma > 2). \quad (17)$$

となる. このように次数分布はスケールフリーの性質を満たし, その冪指数は結合の生じやすさの冪指数と一致している.

一方 $\gamma < 2$ の場合を同様に計算すると

$$\frac{5-2\gamma}{2-\gamma} < k < \frac{2mN}{\ln N} + 2, \quad (18)$$

で

$$P(k) \simeq \frac{2m}{\ln N} k^{-2} \quad (1 < \gamma < 2). \quad (19)$$

となることがわかる. この場合は次数分布の冪指数は γ によらず常に 2 となっている. 次数分布はノードが埋め込まれた空間の次元 d にはよらない.

4 クラスタリング係数

ノードの結合の生じやすさ a の関数としてのクラスタリング係数は

$$C(a) = \frac{\int r_3(a, a', a'') \rho(a') \rho(a'') da' da''}{\int r(a, a') r(a, a'') \rho(a') \rho(a'') da' da''} \quad (20)$$

のように計算される．ここで $r_3(a, a', a'')$ は，結合の生じやすさがそれぞれ a, a', a'' となる3つのノードが互いにつながって三角形を構成する確率である． $a < a' < a''$ の場合 $b = a^{1/d}, b' = a'^{1/d}, b'' = a''^{1/d}$ とおくと最大値ノルムの性質から面倒な計算を経て

$$r_3(a, a', a'') = \begin{cases} r(a, a')r(a, a'') & ((\theta a' a'' > 1) \text{ or } (\frac{1}{b} \geq \frac{1}{b'} + \frac{1}{b''})) \\ \theta^2 \left(bb' b'' (b + b' + b'') - \frac{b' b'' + b'' b + b b'}{\theta^{1/d}} + \frac{1}{\theta^{2/d}} \right)^d & ((\theta a' a'' < 1) \text{ and } (\theta^{1/d} (b' b'' + b'' b + b b') > 2)) \\ \theta^2 (2bb' b'' (b + b' + b'') - b' b'' - b'' b - b b')^d / 4^d & (\text{otherwise}) \end{cases}$$

と求まる． $\gamma < 2$ では積分範囲の大きいほう ($a \sim N^{1/(\gamma-1)}$) が支配的であり，この領域で $r_3(a, a', a'') \sim r(a, a')r(a, a'')$ とみなせるので $C(a) \sim 1$ と予想できる．実際 a が $1/\sqrt{\theta}$ より小さい場合で $C(a) \sim 1$ となっている．このため次元 d にかかわらず $\gamma < 2$ ではクラスタリング係数は大きい値を保っている．一方， $\gamma > 2$ の場合に関してクラスタリング係数を解析的に求めることは難しい．ここでは平均クラスタリング係数の数値計算による結果を図2(b)に示す．

5 平均ノード間距離

図2に平均ノード間距離の数値計算の結果を示した． $\gamma > 3$ の場合に平均ノード間距離 L は冪則 $L \propto N^\alpha$ に従っている．このような冪則は，格子状のネットワークで一般的に見られるが，ここでの冪指数は $1/d$ よりやや小さいように見える．この原因は明らかではない． $\gamma \sim 3$ の場合に対数的な振る舞い $L \propto \ln N$ をみせ， $\gamma < 3$ では対数よりゆっくりと増加している．このような $\gamma < 3$ での遅い増加は，結合の生じやすさが大きいペア同士は空間上の距離に関係なく繋がっていることに由来する．このことは， $\max(\theta a_i a_j) = \theta N^{\frac{2}{\gamma}-1} \gg 1$ から明らかである．このような空間上の距離によらない結合はショートカットと呼ばれ，ショートカットはネットワークを著しく小さいものにすることが知られており [4]，この結果はもっともである．

6 結語

空間的に埋め込んだネットワークモデルを用いてノードの属性とネットワーク構造の関係を調べた．属性値の分布の冪指数 γ にかかわらず，次数分布の冪指数 μ が2より小さくならないことを示した．次数分布の冪指数が2に近い場合は空間構造が無くてもクラスタリング係数が高くすることができる．一方，次数分布の冪指数が2から離れている場合には空間構造の次元がクラスタリング係数を決めている．さらに次数分布の冪指数が3を超える場合，ネットワークはスモールワールドとしての特性（平均ノード間距離が小さい）を失う．

多くの実在のネットワークで次数分布の冪指数 μ は2から3の間である [6, 2, 3]．冪指数が3に近い場合には，高クラスタリング係数を生むために何らかの空間的な構造が関わっていることが予想される．一方，冪指数が2に近ければ空間構造は高クラスタリング

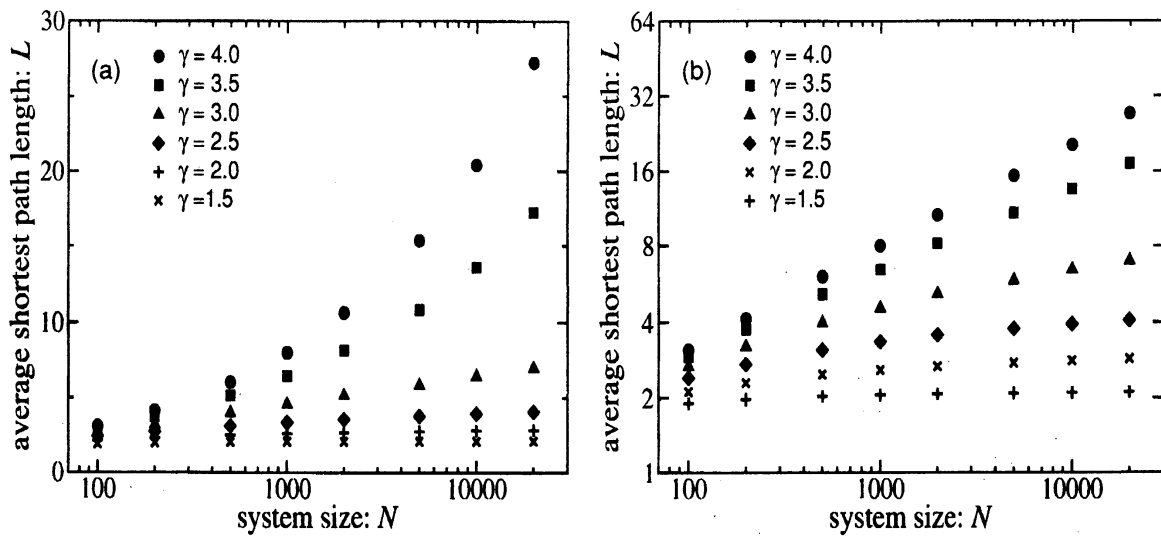


図 2: (a) 平均ノード間距離 L 対ノード数の片対数プロット. パラメータは $d = 2$ で上から下へ $\gamma = 4.0, 3.5, 3.0, 2.5, 2.0, 1.5$ である. (b) 同じデータに対する両対数プロット

係数の再現に必ずしも必要ではない. 本研究で用いたモデルをさらに解析すると $\mu < 3$ では次数相関が負であることもいえる [10]. 人間関係や生物に見られるネットワークでは次数相関が正であることが知られており, モデルの結果と一致しない. この点についてはモデルの一般化を含めた更なる研究が必要である.

- [1] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [2] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002); *Evolution of Networks* (Oxford University Press, New York, 2003).
- [3] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
- [4] R. Cohen and S. Havlin, *Phys. Rev. Lett.* **90**, 058701 (2003).
- [5] S. H. Strogatz, *Nature* **410**, 268 (2001).
- [6] R. Albert and A. -L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [7] A. -L. Barabási and R. Albert, *Science* **286**, 509 (1999)
- [8] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, *Proc. Nat. Acad. Sc. USA* **97**, 11149 (2000).
- [9] N. Masuda, H. Miwa, N. Konno, *Phys. Rev. E* **70**, 036124 (2004); N. Masuda, H. Miwa, N. Konno, *Phys. Rev. E* **71**, 036108 (2005).
- [10] S. Morita, *Phys. Rev. E* **73**, 035104R (2006).